[9] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[10] S. R. Kosaraju, "Efficient string matching," unpublished.
[11] L. Prasad and S. Iyengar, "High performance algorithms for object recognition problem by multiresolution template matching," in *Proc. 1995 IEEE 7th Int. Conf. Tools Artificial Intelligence*, Herndon, VA, Nov. 1995, pp. 362–365.
[12] P. Remagnino, P. Brand, and R. Mohr, "Correlation techniques in adaptive template matching with uncalibrated cameras," *Proc. SPIE*, vol. 2356, pp. 252–263, 1995.
[13] F. Saitoh, "Template searching in a semiconductor chip image using partial random search and adaptable search," *J. Jpn. Soc. Precision Eng.*, vol. 61, no. 11, pp. 1604–1608, Nov. 1995.
[14] H. Senoussi and A. Saoudi, "Quadtree algorithm for template matching on a pyramid computer," *Theoret. Comput. Sci.*, vol. 136, no. 2, pp. 387–417, Dec. 1994.

On Kalman Filter Solution of Space–Time Interpolation

Toshio M. Chin

***Abstract*—The approximate Kalman filtering algorithm presented in [1] for image sequence processing can introduce unacceptable negative eigenvalues in the information matrix and can have degraded performance in some applications. The improved algorithm presented in this note guarantees a positive definite information matrix, leading to more stable filter performance.**

***Index Terms*—Data assimilation, Kalman filter, Markov random field, recursive least-squares, satellite imaging.**

I. Introduction

In [1], an approximate Kalman filtering method has been introduced for time-recursive solution of an image sequence reconstruction/restoration problem. Let the vector $\mathbf{x}_k$ be the collection of the unknowns over an image grid with N pixels at a time-index k, and $\mathbf{y}_k$ be the under-constraining (e.g., sparse) observations of $\mathbf{x}_k$. The solution is then sought for the time-varying, space-time optimization (recursive least-squares [2]) problem

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_k} \sum_{i=1}^{k} \{\|\mathbf{x}_i - \mathbf{F}_i\mathbf{x}_{i-1}\|^2_{\mathbf{M}_i} + \|\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i\|^2_{\mathbf{N}_i}\} \quad (1)$$

for $\mathbf{M}_1 = \mathbf{0}$, where $\|A\|^2_B \equiv A^T B A$ denotes the quadratic norm of A with a positive definite weight matrix B. The minimizing solution $\hat{\mathbf{x}}_k$ for (1) can be computed time-recursively by applying a Kalman filter algorithm to the dynamic system

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{w}_k \quad (2)$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \quad (3)$$

Manuscript received February 13, 2000; revised December 15, 2000. This work was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, under a contract with the National Aeronautics and Space Administration and in part at Rosenstiel School of Marine and Atmospheric Science, University of Miami, Coral Gables, FL, through the Office of Naval Research under Grant N00014-95-1-0257. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael R. Frater.

The author is with Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (e-mail: mike.chin@jpl.nasa.gov).

Publisher Item Identifier S 1057-7149(01)02469-1.

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually uncorrelated sequence of zero-mean Gaussian processes with covariances $\mathbf{M}^{-1}$ and $\mathbf{N}^{-1}$. A key aspect in this formulation is that images are reconstructed/restored by optimizing their local attributes such as spatial gradients against closeness to the observations. The algebraic manifestation of this locality is that the given matrices $\mathbf{F}, \mathbf{H}, \mathbf{M}, \mathbf{N}$ (denoted without the time-index k for lack of ambiguity hereafter) have sparsely banded structures. Examples include the partial differential equations of fluid dynamics used for $\mathbf{F}$ [3] and smoothness constraints of low-level vision used for $\mathbf{H}$ [1].

There is an economical incentive to exploit the given sparse matrix structures during recursive computation of $\hat{\mathbf{x}}_k$, as the Kalman filter requires co-recursion of an $N \times N$ matrix for a typically large N. For example, N can range $10^4 \sim 10^7$ in geophysical applications known as "data assimilation" (dynamic mapping of atmospheric and oceanic variables from sparse observations including satellite measurements [3], [4]). In standard Kalman filter, the covariance matrix $\hat{\mathbf{P}}_k$ of the estimation error $\mathbf{x}_k - \hat{\mathbf{x}}_k$ is recursively computed. The algorithm presented in [1] performs recursion of the estimation error information matrix $\hat{\mathbf{L}}_k \equiv \hat{\mathbf{P}}_k^{-1}$ instead, as the inherent sparseness in the formulation is more apparent in the structure of the information matrix (e.g., $\mathbf{M}$ and $\mathbf{N}$ themselves are the information matrices for $\mathbf{w}_k$ and $\mathbf{v}_k$, respectively). By limiting the matrix bandwidth appropriately, the information matrix $\hat{\mathbf{L}}_k$ can approximate the error covariances compactly using only $O(N)$ nonzero elements. In fact, the elements of the information matrix can be identified as the parameters of a Markov random field (or regression in space) [1].

The algorithm presented in [1] for recursion of a sparsely approximated $\hat{\mathbf{L}}_k$, however, does not guarantee positive definiteness. Negative eigenvalues are not only infeasible in an information matrix but also causes of numerical inaccuracy and inefficiency (e.g., during iterative inversion of $\hat{\mathbf{L}}_k$). This note presents an alternative recursion scheme that preserves positive definiteness in the approximated $\hat{\mathbf{L}}_k$. Also, the new scheme provides a measure of accuracy of approximation.

II. Recursion of Information Matrix

The Kalman filter equations for recursion of the optimal estimate $\hat{\mathbf{x}}_k$ based on the dynamic system (2)–(3) can be written as

$$\overline{\mathbf{x}}_k = \mathbf{F}\hat{\mathbf{x}}_{k-1} \quad (4)$$

$$\hat{\mathbf{L}}_k(\hat{\mathbf{x}}_k - \overline{\mathbf{x}}_k) = \mathbf{H}^T\mathbf{N}(\mathbf{y}_k - \mathbf{H}\overline{\mathbf{x}}_k) \quad (5)$$

where $\hat{\mathbf{L}}_k$ is the information matrix associated with the estimation error $\mathbf{x}_k - \hat{\mathbf{x}}_k$. We also denote as $\overline{\mathbf{L}}_k$ the information matrix corresponding to the prediction error $\mathbf{x}_k - \overline{\mathbf{x}}_k$. The sequence $\hat{\mathbf{L}}_k$ associated with the optimal estimates can be obtained by minimizing the "information" in a manner consistent with the maximum entropy principle [5]. One way to realize this is through the following recursion.

Theorem (Recursion of information matrix): Assume that the matrix $\hat{\mathbf{L}}_{k-1} + \mathbf{F}^T\mathbf{M}\mathbf{F}$ is symmetric and strictly positive definite. The estimation error information matrix $\hat{\mathbf{L}}_k$ in (5) can then be obtained by the recursion

$$\mathbf{L}_k(\boldsymbol{\Theta}_k) = \|\boldsymbol{\Theta}_k\|^2_{\hat{\mathbf{L}}_{k-1}} + \|\mathbf{I} - \mathbf{F}\boldsymbol{\Theta}_k\|^2_{\mathbf{M}} \quad (6)$$

$$\hat{\mathbf{L}}_k = \mathbf{L}_k(\boldsymbol{\Theta}_k) + \mathbf{H}^T\mathbf{N}\mathbf{H} \quad (7)$$

if $\boldsymbol{\Theta}_k$ in (6) is chosen to minimize the trace of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ for each k. Moreover, $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ with the minimum trace is unique and is equal to $\overline{\mathbf{L}}_k$.

The recursions (4)–(7) are initialized as $\hat{\mathbf{x}}_0 = \mathbf{0}$ and $\hat{\mathbf{L}}_0 = \mathbf{0}$ for computation of the minimizing $\hat{\mathbf{x}}_k$ in (1). For most applications

$\mathbf{F}^T\mathbf{M}\mathbf{F}$ is positive definite; hence so is $\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F}$. Positive definiteness and symmetry of $\hat{\mathbf{L}}_k$ can be easily verified (and imposed numerically) from (6) and (7), even for an arbitrary $\boldsymbol{\Theta}_k$. A systematic approximation of $\hat{\mathbf{L}}_k$ can be performed by constraining the free parameter $\boldsymbol{\Theta}_k$ so that $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ would have a desired sparse structure. Such an approach is explored in Section IV.

To prove the theorem the optimal expressions for $\boldsymbol{\Theta}_k$ and $\overline{\mathbf{L}}_k$ are derived first. Minimization of the trace of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ is equivalent to minimizing each of the (positive real) diagonal element

$$\ell_{jj} = \|\boldsymbol{\theta}_j\|^2_{\hat{\mathbf{L}}_{k-1}} + \|\mathbf{u}_j - \mathbf{F}\boldsymbol{\theta}_j\|^2_{\mathbf{M}} \tag{8}$$

$j = 1, 2, \cdots, N$, where $\mathbf{u}_j$ and $\boldsymbol{\theta}_j$ are the jth columns of $\mathbf{I}$ and $\boldsymbol{\Theta}_k$, respectively. Minimizing individually with respect to $\boldsymbol{\theta}_j$ results in the normal equation $(\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F})\,\boldsymbol{\theta}_j = \mathbf{F}^T\mathbf{M}\mathbf{u}_j$, which leads to the optimal $\boldsymbol{\Theta}_k$ as

$$\overline{\boldsymbol{\Theta}}_k = (\mathbf{F}^T\mathbf{M}\mathbf{F} + \hat{\mathbf{L}}_{k-1})^{-1}\mathbf{F}^T\mathbf{M}. \tag{9}$$

The optimum is unique since $\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F}$ is nonsingular. The prediction error information matrix then becomes

$$\begin{aligned}\overline{\mathbf{L}}_k &= \mathbf{L}_k(\overline{\boldsymbol{\Theta}}_k)\\ &= \mathbf{M} + \overline{\boldsymbol{\Theta}}_k{}^T(\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F})\overline{\boldsymbol{\Theta}}_k\\ &\quad - \overline{\boldsymbol{\Theta}}_k{}^T\mathbf{F}^T\mathbf{M} - \mathbf{M}\mathbf{F}\overline{\boldsymbol{\Theta}}_k\\ &= \mathbf{M} + \overline{\boldsymbol{\Theta}}_k\mathbf{F}^T\mathbf{M} - \overline{\boldsymbol{\Theta}}_k{}^T\mathbf{F}^T\mathbf{M} - \mathbf{M}\mathbf{F}\overline{\boldsymbol{\Theta}}_k\\ &= \mathbf{M} - \mathbf{M}\mathbf{F}(\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{M}\end{aligned} \tag{10}$$

which turns (7) to $\hat{\mathbf{L}}_k = \overline{\mathbf{L}}_k + \mathbf{H}^T\mathbf{N}\mathbf{H}$. We next show that these expressions for the information matrices are indeed consistent with the Kalman filter error covariances.

The Kalman filter equations can be derived using the maximum likelihood principle [6]. Given a generic observation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$ where $\mathbf{v}$ is a zero-mean random vector with covariance $\mathbf{N}^{-1}$, the maximum likelihood (ML) estimate can be given as $\mathbf{x}^{(\mathrm{ML})} = \mathbf{P}\mathbf{H}^T\mathbf{N}\mathbf{y}$, while the associated estimation error covariance $\mathbf{P}$ and information $\mathbf{L}$ matrices are $\mathbf{P} = \mathbf{L}^{-1}$ and $\mathbf{L} = \mathbf{H}^T\mathbf{N}\mathbf{H}$, respectively [5]. These formulae are used in the derivation below.

We first combine the dynamic equation (2) with the estimate $\hat{\mathbf{x}}_{k-1}$ and $\hat{\mathbf{L}}_{k-1}$ given from the previous time step to derive the expression for the prediction $\overline{\mathbf{x}}_k$ and $\overline{\mathbf{L}}_k$. The square root of the information matrix is a whitening operator for the corresponding estimation error. If $\hat{\boldsymbol{\Gamma}}$ is a square root such that $\hat{\boldsymbol{\Gamma}}^T\hat{\boldsymbol{\Gamma}} \equiv \hat{\mathbf{L}}_{k-1}$, then

$$\hat{\boldsymbol{\Gamma}}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) = \boldsymbol{\delta}_{k-1} \tag{11}$$

where $\boldsymbol{\delta}_{k-1}$ is a zero-mean process with covariance matrix of identity $\mathbf{I}$. Combining this equation with (2) would yield

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}\hat{\mathbf{x}}_{k-1}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{\Gamma}} & \mathbf{0}\\ -\mathbf{F} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{k-1}\\ \mathbf{x}_k\end{bmatrix} + \begin{bmatrix}\boldsymbol{\delta}_{k-1}\\ -\mathbf{w}_k\end{bmatrix} \tag{12}$$

where $\boldsymbol{\delta}_{k-1}$ and $\mathbf{w}_k$ are mutually uncorrelated. Applying the ML estimation formula to (12) would result in

$$\begin{bmatrix}\mathbf{x}^{(\mathrm{ML})}_{k-1}\\ \mathbf{x}^{(\mathrm{ML})}_{k}\end{bmatrix} = \mathcal{P}\begin{bmatrix}\hat{\mathbf{L}}_{k-1}\hat{\mathbf{x}}_{k-1}\\ \mathbf{0}\end{bmatrix} \tag{13}$$

and

$$\mathcal{P} = \begin{bmatrix}\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F} & -\mathbf{F}^T\mathbf{M}\\ -\mathbf{M}\mathbf{F} & \mathbf{M}\end{bmatrix}^{-1} \equiv \begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{bmatrix} \tag{14}$$

where $\mathbf{P}_{ij}$ denote the $N \times N$ marginal error covariance matrices. The one-step ahead prediction $\overline{\mathbf{x}}_k$ and its error covariance are given by $\mathbf{x}^{(\mathrm{ML})}_k$ and $\mathbf{P}_{22}$, respectively. The desired expressions can then be obtained by applying the matrix inversion lemma [5] as follows:

$$\begin{aligned}\overline{\mathbf{x}}_k &= \mathbf{P}_{21}\hat{\mathbf{L}}_{k-1}\hat{\mathbf{x}}_{k-1}\\ &= \mathbf{M}^{-1}\mathbf{M}\mathbf{F}\mathbf{P}_{11}\hat{\mathbf{L}}_{k-1}\hat{\mathbf{x}}_{k-1}\\ &= \mathbf{F}(\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F} - \mathbf{F}^T\mathbf{M}\mathbf{M}^{-1}\mathbf{M}\mathbf{F})^{-1}\hat{\mathbf{L}}_{k-1}\hat{\mathbf{x}}_{k-1}\\ &= \mathbf{F}\hat{\mathbf{x}}_{k-1}\end{aligned} \tag{15}$$

which is (4) and

$$\begin{aligned}\overline{\mathbf{L}}_k &= \mathbf{P}_{22}^{-1}\\ &= \mathbf{M} - \mathbf{M}\mathbf{F}(\hat{\mathbf{L}}_{k-1}+\mathbf{F}^T\mathbf{M}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{M}\end{aligned} \tag{16}$$

which is (10).

We now combine the prediction $\overline{\mathbf{x}}_k$ and $\overline{\mathbf{L}}_k$ with the observation equation (3) to obtain an ML estimate. Again, a square root $\overline{\boldsymbol{\Gamma}}$ of $\overline{\mathbf{L}}_k$, such that $\overline{\boldsymbol{\Gamma}}^T\overline{\boldsymbol{\Gamma}} = \overline{\mathbf{L}}_k$, is a whitening operator for the prediction error

$$\overline{\boldsymbol{\Gamma}}(\mathbf{x}_{k-1} - \overline{\mathbf{x}}_{k-1}) = \boldsymbol{\delta}_k \tag{17}$$

where $\boldsymbol{\delta}_k$ has zero-mean and a covariance of $\mathbf{I}$. Combining with (3) results in

$$\begin{bmatrix}\overline{\boldsymbol{\Gamma}}\overline{\mathbf{x}}_k\\ \mathbf{y}_k\end{bmatrix} = \begin{bmatrix}\overline{\boldsymbol{\Gamma}}\\ \mathbf{H}\end{bmatrix}\mathbf{x}_k + \begin{bmatrix}\boldsymbol{\delta}_k\\ \mathbf{v}_k\end{bmatrix} \tag{18}$$

from which the ML estimate would yield the filtered estimate $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{L}}_k$. Noting that $\boldsymbol{\delta}_k$ and $\mathbf{v}_k$ are uncorrelated, the ML estimation formula would yield

$$\hat{\mathbf{L}}_k = \overline{\mathbf{L}}_k + \mathbf{H}^T\mathbf{N}\mathbf{H} \tag{19}$$

which is the optimal version of (7), and

$$\hat{\mathbf{L}}_k\hat{\mathbf{x}}_k = \overline{\mathbf{L}}_k\overline{\mathbf{x}}_k + \mathbf{H}^T\mathbf{N}\mathbf{y}_k \tag{20}$$

which easily leads to (5).

III. Interpretations

The theorem specifies an optimal strategy to choose $\overline{\mathbf{L}}_k$ among the candidate matrices $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ given by (6). To gain insights into these candidate matrices, let $\boldsymbol{\xi}_k \equiv \mathbf{F}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})$ be the forecast error whose information matrix is denoted as $\hat{\mathbf{L}}_k \equiv (\mathbf{F}\hat{\mathbf{P}}_k\mathbf{F}^T)^{-1}$. The one-step ahead prediction error $\overline{\mathbf{x}}'_k \equiv \mathbf{x}_k - \overline{\mathbf{x}}_k$ can then be written as $\overline{\mathbf{x}}'_k = \boldsymbol{\xi}_k + \mathbf{w}_k$. We now consider expanding the prediction error as $\overline{\mathbf{x}}'_k = (\mathbf{I} - \boldsymbol{\Phi}_k)\overline{\mathbf{x}}'_k + \boldsymbol{\Phi}_k\overline{\mathbf{x}}'_k$ using a free matrix parameter $\boldsymbol{\Phi}_k$. We then assign $\boldsymbol{\xi}_k = (\mathbf{I} - \boldsymbol{\Phi}_k)\overline{\mathbf{x}}'_k$ and $\mathbf{w}_k = \boldsymbol{\Phi}_k\overline{\mathbf{x}}'_k$, or equivalently

$$\begin{bmatrix}0\\ 0\end{bmatrix} = \begin{bmatrix}\mathbf{I} - \boldsymbol{\Phi}_k\\ \boldsymbol{\Phi}_k\end{bmatrix}\overline{\mathbf{x}}'_k + \begin{bmatrix}-\boldsymbol{\xi}_k\\ -\mathbf{w}_k\end{bmatrix} \tag{21}$$

from which the (posterior) information matrix for $\overline{\mathbf{x}}'_k$ can be obtained from the ML formula as

$$\mathbf{L}_k(\boldsymbol{\Phi}_k) = (\mathbf{I} - \boldsymbol{\Phi}_k)^T\tilde{\mathbf{L}}_k(\mathbf{I} - \boldsymbol{\Phi}_k) + \boldsymbol{\Phi}_k^T\mathbf{M}\boldsymbol{\Phi}_k. \tag{22}$$

By re-parameterization $\boldsymbol{\Phi}_k \equiv \mathbf{I} - \mathbf{F}\boldsymbol{\Theta}_k$, the set of matrices defined as (22) can be shown identical to the set of the candidate matrices of (6). Thus, the candidates for the prediction error information matrix can be

interpreted as the results of the expansion (21), which partitions $\overline{\mathbf{x}}'_k$ into components based on the forecast error ($\boldsymbol{\xi}_k$) and dynamics error ($\mathbf{w}_k$). The partitioning of $\overline{\mathbf{x}}'_k$ allows access to the matrix structure (for sparse approximations) through its free parameters. It is reasonable from an information theoretic perspective [5] that the optimal choice $\overline{\mathbf{L}}_k$ happens to be the candidate with the least information (or maximum uncertainty), measured by the trace (sum of eigenvalues). The trace of the candidate matrices can thus serve as a measure of optimality in this sense. Note, however, that the trace is not an explicit indicator for other matrix properties, such as the distribution of eigenvalues and structures of the eigenvectors, that may also affect filter performance when $\overline{\mathbf{L}}_k$ is approximated.

IV. Approximations

Given sparsely banded matrices $\mathbf{F}, \mathbf{H}, \mathbf{M}, \mathbf{N}$ (for all k) and $\hat{\mathbf{L}}_0$, the recursive equations (16) and (19) cannot maintain a sparse structure in $\overline{\mathbf{L}}_k$ and $\hat{\mathbf{L}}_k$ because of the prediction step (16). In [1], the matrix inverse in (16) is approximated by a series expansion based on the Jacobi iteration

$$(\hat{\mathbf{L}}_{k-1} + \mathbf{F}^T\mathbf{M}\mathbf{F})^{-1} = \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega}\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega}\boldsymbol{\Lambda}^{-1} - \cdots \quad (23)$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are identical to those of the matrix $\hat{\mathbf{L}}_{k-1} + \mathbf{F}^T\mathbf{M}\mathbf{F}$ while $\boldsymbol{\Omega} \equiv \hat{\mathbf{L}}_{k-1} + \mathbf{F}^T\mathbf{M}\mathbf{F} - \boldsymbol{\Lambda}$. The bandwidth of the approximated inverse increases as more terms participate in the series. Specifically, if $\mathbf{F}$ is diagonal for all k then the first two terms in the right hand side of (23) can be used to contain the matrix bandwidth in the information matrix in (16). If $\mathbf{F}$ has nonzero off-diagonal elements, only the first term in the series can be used to prevent spreads in matrix bandwidth in general. This approach to information matrix approximation is hereafter referred to as the *series truncation* method. A major drawback of this approximation method is that negative eigenvalues can be introduced into the information matrix.

The alternative recursion (6)–(7) introduced in this note can maintain positive definitiveness in the information matrix. As a direct application of the theorem presented earlier, we can impose an arbitrary sparse structure on the information matrix by minimizing its trace while constraining the appropriate elements of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ to equal to zero. Numerically this leads to a minimization of a quadratic cost with many nonlinear constrain equations, which can be optimized iteratively [7]. We refer to this approximation approach as the *direct* method.

Constrained minimization in the direct method can be computationally intensive. An economical alternative is to restrict $\boldsymbol{\Theta}_k$ to be a diagonal matrix. This, referred to as the *diagonal* method, tends to preserve the bandwidth in $\hat{\mathbf{L}}_k$. To compute the optimal diagonal $\boldsymbol{\Theta}_k$, we let $\boldsymbol{\theta}_j = \theta_{jj}\mathbf{u}_j$ and minimize (8) with respect to each diagonal element θ_{jj}, resulting in

$$\text{diag}\,(\boldsymbol{\Theta}_k) = \text{diag}\,(\mathbf{F}^T\mathbf{M})/\text{diag}\,(\hat{\mathbf{L}}_{k-1} + \mathbf{F}^T\mathbf{M}\mathbf{F}). \quad (24)$$

The set of feasible $\boldsymbol{\Theta}_k$ in the diagonal method is a subset of the feasible set in the direct method. The trace of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ resulting from the diagonal method would then be larger than or equal to that resulting from the direct method. In this context the direct method can yield the best parameters of the Markov random field (as which a sparse information matrix can be interpreted [1], [3]) approximating the estimation error at each time step of Kalman filter.

V. Examples

For comparison, the approximation methods are applied to a one-dimensional (1-D) version of the image sequence reconstruction problem in [1]

$$\min_{x(s,t)} \left[\int_0^{k\tau} \int_{\mathcal{D}} w_1 \left\| \frac{\partial x}{\partial t} \right\|^2 + w_2 \left\| \frac{\partial x}{\partial s} \right\|^2 ds\, dt + \sum_{i=1}^{k} \|y_i - x(s_i, i\tau)\|^2 \right] \quad (25)$$

where the unknown $x(s,t)$ is defined in part over a 1-D cyclic spatial domain $\mathcal{D}$ and is observed as y_k at a single randomly selected location s_k at time $t = k\tau$ for a fixed interval τ. For formulation of a discrete Kalman filter, the first order differences are used for the derivatives, and the second and third quadratic terms in (25) are lumped into a single, vector equation (3) (see [1]). The parameters used here are $w_1 = 1$, $w_2 = 0.001$, and $N = 32$.

A unit-magnitude sine wave has been reconstructed from its noiseless (but very sparse) observations y_k using the optimal and approximated Kalman filter algorithms. The information matrix was approximated by constraining it to be cyclic tri-diagonal, and the approximation was executed with each of the truncated series, direct, and diagonal methods. All three approximation methods have performed well with respect to the optimal (nonapproximated) Kalman filter. The root mean square (rms) difference between the approximate and optimal estimates was 0.012 at $k = 32$ for each of the three approximation methods. For the direct and diagonal methods, the values of the trace of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ were both 0.8% higher than that of (the optimal) $\overline{\mathbf{L}}_k$, on the average over 100 time steps. The small discrepancy in the trace values (between the optimal and approximated) is consistent with the low-rms error in approximations. For the series truncation method, no negative eigenvalue was observed during the first 100 time steps.

Next, a translating version of the unit-magnitude sine wave is reconstructed. This problem is motivated by the data assimilation applications [4] in which estimation of waves in motion from sparse measurements is often a key task. To incorporate a known translation speed c, we replace the first integrand in (25) with

$$w_1 \left\| \frac{\partial x}{\partial t} - c\frac{\partial x}{\partial s} \right\|^2$$

which is discretized to a cyclically tridiagonal state-transition matrix $\mathbf{F}$ (by a Lax-Wendroff scheme). In general, $\mathbf{F}$ here would have a more complex structure than in the previous case without translation. With $c = 0.1$ and keeping all other reconstruction parameter values identical to the previous case, the optimal and approximate filters have been applied to the sparse observations y_k. The information matrix was approximated to be cyclic penta-diagonal. The rms differences between the approximate and optimal estimates at $k = 32$ were 10.200, 0.030, and 0.049 for the series truncation, direct, and diagonal methods, respectively. Along with the unacceptably high-rms value, negative eigenvalues in the approximated information matrix have been observed (especially in early time steps) for the series truncation method, indicating that this approximation method is unsuitable for the problem. For the direct and diagonal methods, the respective trace values of $\mathbf{L}_k(\boldsymbol{\Theta}_k)$ were 11.1% and 30.4% higher than the trace of the optimal $\overline{\mathbf{L}}_k$ (averaged over 100 time steps). The relative increase in trace is consistent with the relative rms accuracy for these two approximation methods.

Clearly, the alternative recursion presented in this note, with its ability to impose positive definiteness in the approximated information

matrix, improves consistency in filter performance dramatically. The best sparse approximation given a desired matrix structure can be computed with the direct method, useful for predicting (through a miniaturized test case like those considered in this section) the best possible performance of an approximated Kalman filter.

REFERENCES

[1] T. M. Chin, W. C. Karl, and A. S. Willsky, "Probabilistic and sequential computation of optical flow using temporal coherence," *IEEE Trans. Image Processing*, vol. 3, pp. 773–788, Nov. 1994.

[2] H. W. Sorenson, "Least-squares estimation: From Gauss to Kalman," *IEEE Spectrum*, vol. 7, pp. 63–68, Jan. 1970.

[3] A. Asif and J. M. F. Moura, "Data assimilation in large time-varying multidimensional fields," *IEEE Trans. Image Processing*, vol. 8, pp. 1593–1607, Nov. 1999.

[4] M. Ghil and P. Malanotte-Rizzoli, "Data assimilation in meteorology and oceanography," *Adv. Geophys.*, vol. 33, pp. 141–266, 1991.

[5] D. E. Catlin, *Estimation, Control, and the Discrete Kalman Filter*. New York: Springer-Verlag, 1989.

[6] R. Nikoukhah, A. S. Willsky, and B. C. Levy, "Kalman filtering and Riccati equations for descriptor systems," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 1325–1342, Sept. 1992.

[7] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming*. New York: Wiley, 1979.